Renewable and Sustainable Energy Reviews 39 (2014) 1143–1154

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

Quantizing the deterministic nonlinearity in wind speed time series

Haidar Samet*, Fatemeh Marzbani

School of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran

A R T I C L E I N F O

Article history:
Received 14 December 2011
Received in revised form
26 January 2014
Accepted 19 July 2014
Available online 9 August 2014

Keywords:
Wind speed
ARMA
Surrogate data
Nonlinear analysis
Markov
Grey
Empirical Mode Decomposition

A B S T R A C T

Linear models are capable of capturing the Linear Deterministic (LD) component of the time series. In order to benefit from both Nonlinear Deterministic (ND) and LD components during the prediction procedure, it is necessary to employ nonlinear models. The complexity of the prediction algorithm increases when nonlinear models are utilized. Hence, before applying nonlinear models the presence of nonlinear component should be confirmed. Although surrogate data technique uses various tests to indicate the nonlinearity, in many cases its test results are different and in conflict with each other. The reason is time series include LD and ND components together and giving a strict answer about nonlinearity cannot be applicable. Here instead of such a strict answer, by quantizing the ND component, a new index (a number between 0 and 1) is proposed (the closer to 1 the more ND components). In this method first we use ARMA models. The residual series is used to calculate the proposed index which it contains all components of the original series except LD. The proposed procedure is applied to three different case studies. Furthermore, the performance of some nonlinear prediction methods (Markov, Grey, Grey–Markov, EMD–Grey, NARnet and ARMAX) is compared with the proposed index.

Contents

1. Introduction . . . 1144
2. The proposed algorithm . . . 1146
3. Test cases . . . 1146
4. Linear ARMA models and residual series . . . 1146
5. Nonlinear analyses . . . 1147
  5.1. Time delay reconstruction . . . 1147
  5.2. Mutual information method . . . 1147
  5.3. Embedding dimension method . . . 1147
  5.4. Surrogate data method . . . 1147
    5.4.1. Third order auto covariance . . . 1148
    5.4.2. Time reversibility . . . 1148
    5.4.3. Standard deviation test (ST) . . . 1148
    5.4.4. Central Frequency Test (CFT) . . . 1148
    5.4.5. Correlation dimension . . . 1148
    5.4.6. Largest Lyapunov Exponent (LLE) . . . 1148
  5.5. Delay vector variance method . . . 1149
6. The proposed quantizing of the nonlinear deterministic component . . . 1150
7. Evaluating some prediction methods with aid of the proposed index . . . 1151
  7.1. Markov chain . . . 1151
  7.2. Grey model . . . 1152
  7.3. Grey–Markov . . . 1152
  7.4. EMD–Grey . . . 1152

* Corresponding author. Tel.: +98 711 6133455; fax: +98 711 2303081.
E-mail address: samet@shirazu.ac.ir (H. Samet).

http://dx.doi.org/10.1016/j.rser.2014.07.130
1364-0321/

7.5. NAR network 1152
7.6. ARMAX. 1152
8. Conclusions 1152
References 1152

1. Introduction

The consumption of electricity is following an increasing manner worldwide. Moreover, the shortage and increasing price of fossil fuels are the main reasons for alternating from conventional sources of energies to renewable energies. Among the family of renewable energies, the wind power is the most promising source. The integration of the wind power to the power grids has increased dramatically. This integration brings new issues to the power system operators and utilities. Such issues stem from the uncertainty of the wind. Hence, it is important to predict the wind speed or power precisely. An accurate wind speed (power) prediction can result in efficient and economic solutions to economic dispatch problems and unit commitment [1].

During the past decades, numerous prediction algorithms have been developed for wind power prediction. The wind power (speed) prediction approaches can be classified into seven classes [2,3]: the persistent method, physical models, statistical methods, learning-based algorithms, Markov models, other techniques, and hybrid approaches. This classification is summarized in Fig. 1.

The first group is the persistent method. This approach is the simplest prediction algorithm. It takes the current measurements as the future value of the time series. The performance of any predictor should be investigated by comparing to the persistent method as the benchmark [3].

Physical models are the second class of this category and initially introduced by McCarthy [4]. Such models provide the future value of a time series using physical inputs such as temperature, pressure, humidity, etc. They are computer based programs solving equations of atmospheric processes [5]. The wind data are traditionally recorded at weather stations and include some constraints, such as representatively and availability problems, high operating cost, and coarse resolution. Therefore, Numeric Weather Prediction (NWP) models are considered as alternative source of the wind data [6]. The NWP models are accurate for short-term prediction and they suffer from the algorithm complexity [7].

The third class is statistical approaches. Such algorithms are based on time series analysis. First, the data pattern is identified using the training measurements. Then, the model residuals are used to tune the obtained parameters. The statistical methods are simple and inexpensive predictors [8,9]. Statistical methods have been used widely for wind prediction [10–12]. Different parameter estimation techniques are used to model wind data such as the Maximum Likelihood Method (MLM), the Method of Moments (MM), the least square method, Weibull function, long normal methods [13,14], and data mining approaches [15,16]. Among the statistical approaches, Auto Regressive Moving Average (ARMA) models [17], Auto Regressive Integrated Moving Average (ARIMA) [18], fractional integrated ARMA models (f-AIRMA) [19] are the most commonly used models. They are time series based algorithms which provide the future values of the wind power time series based on the previous measurements [20].

Grey models are also included in the statistical class of wind predictors. Such models are capable of dealing with systems with poor and uncertain information. The traditional Grey model GM(1,1) is the most frequently used technique among the family of the Grey models. In [21], the annual wind power output is predicted using GM(1,1). The original time series are processed to improve the accuracy of the prediction. The GM(1,1) is used to provide short-term wind power prediction for wind farms in Penghu and Taiwan [22].

The fourth group is the learning-based techniques; the biological neurons were the first inspiration for the Neural Networks (NN). Artificial Neural Network (ANN) [23–28] genetic algorithms [29–31], fuzzy logic [32–36], Support Vector Machines (SVM) [37–41], ANFIS [42], Wavelet transformation [43] are the most frequently learning-based techniques developed for wind power (or speed) prediction.

Markov chain models are the fifth class of wind power predictors. These models are based on transitions from one state to another on a state space. The have been used in a broad range of applications such as wind prediction [44]. In [45] a Markov chain model is proposed for short-term wind power prediction. The states of the chain are constructed based on wind speed, direction and power. A maximum likelihood estimator is utilized to establish the transition matrix. The obtained results affirm the ability of the proposed model to describe the measurements. A first and second order discrete time Markov chain is employed to provide short-term wind power forecast [46]. The obtained results show that the second order of the Markov model enhances the prediction accuracy. Different time horizon wind speed prediction is provided in [47] using a discrete Markov process. The Markov model is capable of describing the stochastic characteristics of the wind speed due to the fact that the future and historical data of the wind speed are independent.

The sixth category is other prediction techniques include spatial correlation models and ensemble predictors. Spatial correlation models benefit from cross correlations between the measurements at the given site and at the neighbor sites [48]. In [49], a spatial wind power prediction is provided using the Augmented Kriging-based Model (AKM). This model is capable of predicting wind power at a new wind farm by using its latitude and longitude. Optimal spatial modeling is developed to locate new wind power resources. Wind Power prediction for 15 min up to 5 h ahead is performed in [50] using spatial correlations and measurements from adjacent wind farms up to 40 km. These inputs are fed to a locally recurrent multilayer network with internal feedback paths. A Latin hypercube sampling-based technique is proposed in [51] to evaluate the Available Transfer Capability (ATC) including huge uncertainty. This technique is used Monte Carlo simulation to determine an efficient system state. Ensemble predictors are another subclass of other techniques. An ensemble short-term predictor is proposed in [52] for wind power prediction. The proposed algorithm is comprised of NN and Gaussian process sub-models. A decision making process is utilized to determine the optimum forecasted values among those obtained from all models. A neural network ensemble method is used for wind power prediction in [53]. In this technique, the ANN is used to synthesize the output of individual networks. The obtained results affirm the outperformance of the proposed algorithm comparing with individual predictors.

Hybrid predictors are the last group of this classification. They refer to a combination of one or more predictors. Such algorithms provide better accuracy comparing to any individual methods. Various hybrid predictors have been developed for wind power or

Nomenclature

x_k	signal value at time k
e_k	unpredictable component
$\sigma^2_{\min\ w}$	minimum target variance for original time series
$\sigma^2_{\min\ e}$	minimum target variance for residual time series
N	number of time series samples
m	number of model parameters
σ^2_a	residual variance
y_k	delay vector at time k
T	embedding delay time
d	embedding dimension
$p(x_n, x_{n+T})$	probability of observing both x_n and x_{n+T}
$p(x_n)$	probability of observingx_n
$I(T)$	information function
$y_{n(i,d)}(d)$	nearest neighbor for $y_i(d)$
τ	time lag
s^x_j	STD value for jth part of the time series
f^x_j	the central frequency for jth part of the time series
θ	heavy side function
D_c	correlation dimension
σ^{*2}	inverse measure of the predictability

speed prediction such as wavelet transformation-SVM [54], Wavelet and NN [55], PSO-ANFIS [56], Wavelet-Particle Swarm Optimization (PSO)–ANFIS [57], ANN-Kalman filtering [58], NN–PSO [59], Grey–Markov[60], Grey–Elman NN [61], Empirical Mode Decomposition (EMD) combined with time series analysis [62], EMD–ANN [63].

Per the review, both the linear and nonlinear methods are proposed and developed for wind speed forecast. The major issue with linear models is only choosing the proper model order and coefficients, while for nonlinear models in addition a decision making process is required to choose the suitable model for the time series under investigation. In this study, a decision making index is proposed. Such an index can be used for two purposes. First, it can be used as a measure to show the nonlinear model prediction improvements over linear models. Second, it can illustrate the feasibility of a specific nonlinear method comparing with other nonlinear techniques.

The proposed index is based on quantizing the nonlinearity of the wind speed time series. Each time series can be composed into three components: linear deterministic (LD), nonlinear deterministic (ND) and unpredictable (U). Deterministic components (LD and ND) contain useful data for the prediction. However, ND components can be used in prediction only by the aid of nonlinear models. The proposed technique quantifies the ratio of ND to summation of LD and ND components which is a number between 0 and 1.

This paper is organized as follows. The proposed quantizing algorithm is described in Section 2. Section 3 provides the details of three case studies used in this research. The linear analysis [64] is explained in Section 4. Section 5 describes the nonlinear

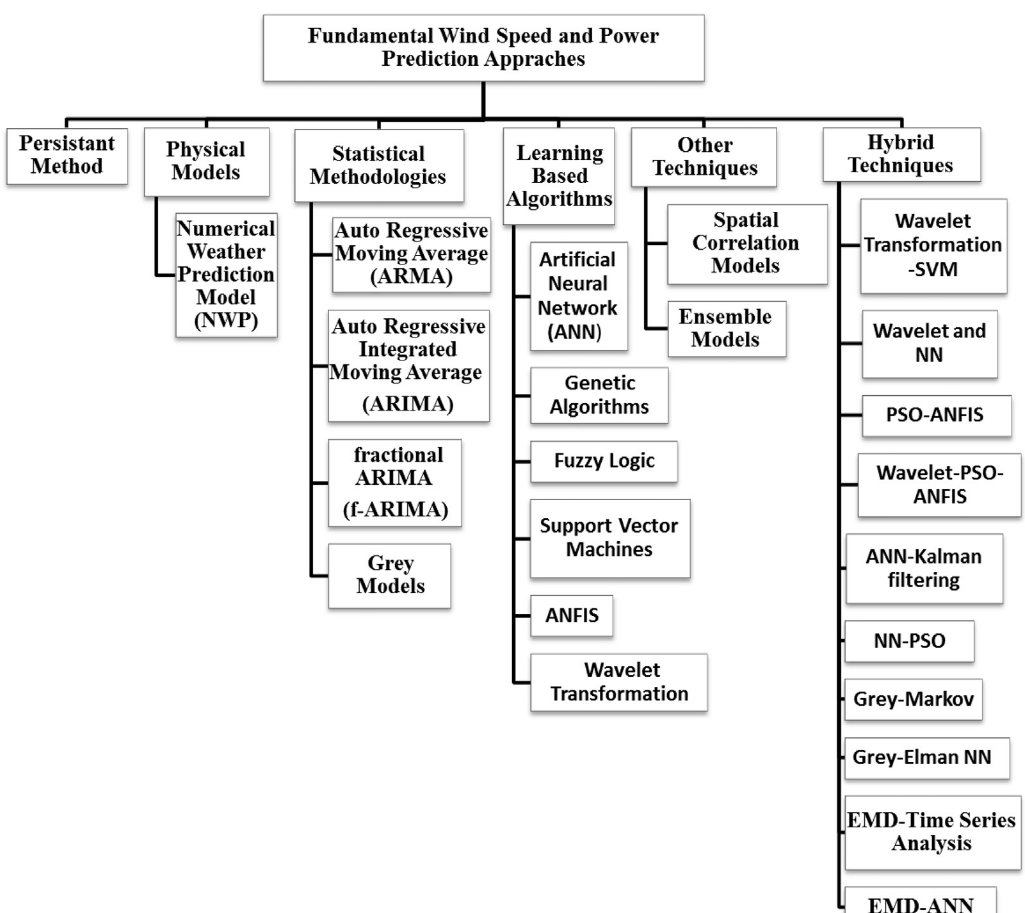

Fig. 1. A classification of wind speed prediction methods.

techniques using time delay reconstruction [65], surrogate data with various test functions [66], and Delay Vector Variance (DVV) [67]. To determine the nonlinearity of time series, the surrogate data method by the aid of different tests is used. It is observed that the outputs of the tests are in conflict with each other. The reason is time series include LD and ND components together and giving a strict answer about nonlinearity cannot be applicable.

In Section 6 the proposed quantizing of the ND component is applied to the test cases. The significance of the ND component over the total determinist component is attained. This new index can answer whether it is appropriate to use nonlinear models for prediction of wind speed or not.

In Section 7, the performance of some nonlinear prediction methods (Markov, Grey, Grey–Markov, EMD–Grey, NARnet and ARMAX) is evaluated using the proposed index. The paper is concluded in Section 8.

2. The proposed algorithm

According to the Wold decomposition theorem, any discrete stationary signal can be decomposed into two uncorrelated components: deterministic (predictable) and nondeterministic components [68]. In this study, the signal x_k (wind speed data) is decomposed as

$$x_k = f(x_{k-1}, x_{k-2,\ldots,}x_{k-p}, e_{k-1}, e_{k-2,\ldots,}e_{k-q}) + e_k \quad (1)$$

where x_k represents the signal value at time k, e_k is the unpredictable component, and the deterministic components are described by $f(.)$ that is a function of a number of previous values and a.

In the present research, we extend the Wold theorem. Here the deterministic components are divided to linear and nonlinear components. Therefore, each signal can be decomposed into three components: LD, ND and unpredictable components. Thus, Eq. (1) can be rewritten as

$$x_k = g(x_{k-1}, x_{k-2,\ldots,}x_{k-p}, e_{k-1}, e_{k-2,\ldots,}e_{k-q}) + h(x_{k-1},$$
$$x_{k-2,\ldots,}x_{k-p}, e_{k-1}, e_{k-2,\ldots,}e_{k-q}) + e_k \quad (2)$$

where the LD, and ND components of the time series is represented by a linear function $g(.)$ and a net nonlinear $h(.)$, respectively. An appropriate ARMA model is capable of capturing LD ($g(.)$ in Eq. (2)) component of the time series. Therefore, the obtained residuals are comprised of only ND and unpredictable components ($h(.)+e_k$).

The proposed quantizing algorithm for the nonlinear deterministic components of time series can be summarized in the following five steps:

1. Consider three components: LD, ND and unpredictable components.
2. Capture the LD components using the proper ARMA models (the residual time series resulted from applying suitable ARMA model can be considered as the time series that only contain the ND and the unpredictable components of the main time series. Hence it is free of the LD component).
3. Use the DVV method described in [67] to obtain the strength of the deterministic components:
   a. Apply DVV algorithm to the original time series and calculate its minimum target variance ($\sigma^2_{\min}$) (this variable is related to the prevalence of the deterministic component over the unpredictable one. In other words, $\sigma^2_{\min}$ yields an inverse measure for the predictability or significance of the deterministic components of the time series).
   b. Apply the DVV method to the residual time series and calculate its minimum target variance.
4. Define a new index called Nonlinearity Ratio (NR) as follows:

$$NR = \left(\frac{\sigma^2_{min\ w}}{\sigma^2_{min\ e}}\right)^{0.5} \times \left(\frac{std_e}{std_w}\right) \quad (3)$$

where $\sigma^2_{min\ w}$ and $\sigma^2_{min\ e}$ are the minimum target variance values for the original and the residual time series, respectively. The standard deviation of the original and the residual time series re represented by std_w and std_e, respectively. Note the deterministic component of the residual signal is equal to nonlinear deterministic component of the original signal. NR lies within the range of 0 and 1. The closer to 1 the more ND components included in the time series. Hence, the efficiency of using nonlinear models for prediction is increased.

3. Test cases

Three test cases are investigated in this study. The first data set is comprised of six time series corresponding to wind speed data recorded every three hours during six years (2003–2008) in Manjil Wind farm (Iran). Manjil is known as a windy city of Iran due to its geographical position in the Alborz mountain chain, a small cleft in Alborz mountain chain that funnels the wind from Manjil to the Qazvin plateau. The largest wind farm of Iran is located near Manjil.

The second case study includes fourteen time series corresponding to the wind speed data recorded every 10 min during 14 months in Soltaniyeh (2007–2008). Soltaniyeh in Zanjan is located in Iran – about 252 km North-West of Tehran, the country's capital. The data is also including the temperature and humidity.

The last data set contains twelve time series recorded during year 2011 at 10 min intervals. The wind speed measurement is performed at the Alberta wind energy system (Canada). The multi turbine wind power curve is used to convert the wind speed to the wind power.

4. Linear ARMA models and residual series

In this section, the proposed algorithm in [69] is accomplished by employing the linear ARMA to obtain residual series. An ARMA

Table 1
The suitable ARMA orders, the first number and second numbers represent p and q in respectively in ARMA(p,q).

Series	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Test case 1														
Order	9,9	9,10	7,9	9,8	9,9	9,9								
Test case 2														
Order	4,5	6,8	8,10	10,6	2,1	8,4	9,7	5,9	4,9	7,5	7,9	8,10	5,10	9,8
Test case 3														
Order	3,5	5,7	4,4	2,2	3,10	4,4	8,8	2,3	2,3	2,3	2,3	8,10		

process of order of (p,q) denoted as *ARMA(p,q)* is given by

$$x_k = \phi_1 x_{k-1} + \phi_2 x_{k-2} + \ldots + \phi_q x_{k-p} + a_k - \theta_1 a_{k-1} - \theta_2 a_{k-2} - \ldots - \theta_q a_{k-q}, \quad k = 1, 2, \ldots \tag{4}$$

where x_k is the value of time series at sample k and a_k is white noise. The time series is assumed to have a zero mean. The time series x_k can be replaced by $x_k - \mu$ when it has a non-zero mean.

The ARMA order estimation includes two phases. First, several model orders are selected as candidates. Second, the corresponding coefficients are estimated the best model is chosen using model adequacy checking methods [64].

In this research, the wind speed time series are analyzed to determine the suitable ARMA models by investigating 121 candidate models

$$\{\text{ARMA}(p,q), \quad p = 0, 1, 2, 3, \ldots, 10, \quad q = 0, 1, 2, 3, \ldots, 10\}$$

This approach is applied to wind speed data using Akaike's information criteria (AIC). In this test, any ARMA model with minimum AIC value is selected as the proper order [64].

$$AIC(m) = \ln \sigma_a^2 + 2m/N \tag{5}$$

where N is the number of time series samples, m is the number of model parameters and σ_a^2 is residual variance.

This test is employed on all time series correspond to the mentioned three test cases and the obtained results are summarized in Table 1.

The obtained residual time series are almost free of LD components and contain only ND and unpredictable components. The residual signals can be used to calculate the ND to Total Deterministic (TD) components ratio.

5. Nonlinear analyses

In this section, time delay reconstruction is employed to obtain their embedding delay time and embedding dimension. Next, several nonlinear measures of surrogate data method are used. Such measures are capable of determining the nonlinear components of the time series under investigation.

5.1. Time delay reconstruction

Time delay reconstruction is a popular nonlinear method. This analysis studies the dynamics of the system using the measured scalar time series. The time delayed time series of the one scalar available series is used by this method. The orbit of the system is traced using multivariate d dimension vectors [70].

$$y_k = [x_k, x_{k+T}, \ldots, x_{k+(d-1)T}] \tag{6}$$

where x_k, T and d are the phase space coordinates, embedding delay time, and embedding dimension, respectively. During the time delay reconstruction procedure, these variables should be determined.

The algorithms for choosing proper values for T and d are explained in the subsequent sub-sections.

5.2. Mutual information method

This method is employed to find the embedding time delay. In order to reduce calculation complexity in time delay reconstruction, it is recommended to minimize the linear and nonlinear correlations between samples of one vector. To achieve this task the information function is described by Abarbanel [71].

$$I(T) = \sum_{n=1}^{N} p(x_n, x_{n+T}) \log_2 \frac{p(x_n, x_{n+T})}{p(x_n) p(x_{n+T})} \tag{7}$$

where $p(x_n, x_{n+T})$ is the probability of observing both x_n andx_{n+T}, $p(x_n)$ is the probability of observing x_n. The proper time delay is the first local minimum of $I(T)$.

This method is applied to all time series. Table 2 shows the time delay for the original and residual time series corresponding to the three test cases. For each test case the rounded average value over its all time series is considered as the proper time delay in our next studies.

5.3. Embedding dimension method

In this study, the TSTOOL package of MATLAB is used to in order to obtain the embedding dimension d [72]. This toolbox minimizes the embedding dimension using the Cao's method [73]. The system is reconstructed by choosing a small value for d

$$y_i(d) = (x_i, x_{i+T}, \ldots, x_{i+(d-1)T}), \quad i = 1, 2, \ldots, N-(d-1)T \tag{8}$$

$A(i,d)$ is defined as below

$$A(i,d) = \frac{\| y_i(d+1) - y_{n(i,d)}(d+1) \|}{\| y_i(d) - y_{n(i,d)}(d) \|} \tag{9}$$

where $\|.\|$ is maximum norm and $y_{n(i,d)}(d)$ is the nearest neighbor for $y_i(d)$.

In the next step $E(d)$ is calculated.

$$E(d) = \frac{1}{N-dT} \sum_{i=1}^{N-dT} A(i,d) \tag{10}$$

The statistical parameters *E1* is defined as below

$$E_1(d) = \frac{E(d+1)}{E(d)} \tag{11}$$

As d increases, this parameter increases and it stops when d is equal to the embedding dimension.

The value of the calculated d for all original and residual time series is shown in Table 3. In this table in addition of considering the rounded average time delays of Table 2, $T=1$ is also considered. Considering $T=1$ is a conservative choice in the context of nonlinearity detection. In the present study, the rounded mean values are used as the proper embedding dimension for each test case.

5.4. Surrogate data method

The surrogate data test is a rigid methodology to investigate the nonlinearity of time series [74]. This method is based on a null hypothesis stating that the available data set has been generated by a linear process [75]. This algorithm is comprised of three major stages. First, a number of surrogate data is generated. The linear properties of such surrogates, such as autocorrelation, mean, variance, etc. are the same as the original data set. Second, some nonlinear measures are applied to the surrogates and the original time series. The obtained values of the utilized nonlinear statistics are compared with each other. In the last stage, the null hypothesis

Table 2
The proper time delay for the original and residual time series.

Series	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Ave.
Test case 1															
Original	5	5	5	5	5	5									5
Residual	3	2	2	4	2	4									2.8
Test case 2															
Original	11	7	8	8	10	4	8	7	9	13	17	8	14	17	10.1
Residual	3	4	2	2	1	2	1	3	1	1	1	2	2	1	1.9
Test case 3															
Original	13	15	18	11	20	12	17	17	14	16	9	16			14.8
Residual	3	2	9	3	2	1	3	3	3	2	1	2			2.8

is rejected when the value of nonlinear measures obtained from the surrogate data is significantly different from those of the original time series.

Several techniques can be used to generate the surrogate time series. In this research, the Iterative Amplitude Adjusted Fourier Transform (iAAFT) is used to generate the surrogate time series [74]. In what follows six nonlinear measures are described.

5.4.1. Third order auto covariance

The third order auto covariance is a generalization of the auto covariance. This measure is a two sided test whose mathematical description is as follows [75]:

$$T_{C3}(\tau)=\frac{\langle x_k x_{k-\tau} x_{k-2\tau}\rangle}{|\langle x_k x_{k-\tau}\rangle|^{3/2}} \tag{12}$$

In which τ is the time lag and the mean function over all samples is depicted by $\langle . \rangle$.

5.4.2. Time reversibility

A stochastic process is said to be time revisable when it is invariant with respect to the reversal of the time scale. The asymmetry due to time reversal is a nonlinear measure described as follows [75]:

$$T_{REV}(\tau)=\frac{\left\langle (x_k-x_{k-\tau})^3\right\rangle}{\left\langle (x_k-x_{k-\tau})^2\right\rangle^{3/2}} \tag{13}$$

5.4.3. Standard deviation test (ST)

In this test, each time series is divided into several sections with the same length and their corresponding Standard Deviation (STD) value is calculated [75].

$$s_j^x=\sqrt{\frac{1}{N^*}\sum_{i=1}^{N^*}(x_i-\overline{x})^2} \tag{14}$$

where s_j^x is the STD value for jth part, $x_i$is ith sample in jth part, $\overline{x}$ is the mean of jth section and N^x is the length of each section. Now the STD value of s_j^x is calculated as the result of this test

$$S^x=\sqrt{\frac{1}{n}\sum_{j=1}^{n}(s_j^x-\bar{s}^x)^2} \tag{15}$$

5.4.4. Central Frequency Test (CFT)

This measure, divides the available time series into some sections with the same length. The central frequency is calculated for each part as the following

$$f_j^x=\frac{2}{N^*}\sum_{i=1}^{N^*/2}\frac{\omega_i x(\omega_i)}{x(\omega_i)} \tag{16}$$

And finally the STD value of f_j^x is calculated by

$$F^x=\sqrt{\frac{1}{n}\sum_{j=1}^{n}(f_j^x-\bar{f}^x)^2} \tag{17}$$

5.4.5. Correlation dimension

In order to calculate the correlation dimension, firstly $C(R)$ (Eq. (18)) is calculated and θ which is a heavy side function will be defined (Eq. (19)). Finally, the correlation dimension is defined by D_c (Eq. (20)) [76].

$$C(R)=\frac{1}{N(N-1)}\sum_{i=0}^{N-dT-1}\sum_{j=0 j\neq i}^{N-dT-1}\theta(R-|y(i)-y(j)|) \tag{18}$$

$$\theta(x)=\begin{cases}0 & ,x<0\\ 1 & ,x\geq 0\end{cases} \tag{19}$$

$$D_c=\lim_{R\to 0}\frac{\log\ C(R)}{\log\ (R)} \tag{20}$$

5.4.6. Largest Lyapunov Exponent (LLE)

The Largest Lyapunov Exponent (LLE) has been frequently used to investigate presence of chaotic behavior as well as nonlinear characteristics of time series as [70]. This measure is based on the divergence of nearby trajectories.

The obtained results from the surrogate data corresponding to three test cases are represented in Tables 4–6. This analysis is performed using different measures by the aid of TSTOOL package. A considerable difference in results obtained from several non-linear measures is observed. The results corresponded to some

Table 3
Embedding dimension for the original and residual time series.

Series	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Ave.
Test case 1															
Original $T=1$	10	9	11	15	13	12									11.7
Original $T=5$	11	11	13	9	8	8									10.3
Residual $T=1$	8	8	9	9	9	9									8.5
Residual $T=3$	8	9	8	8	8	8									8.2
Test case 2															
Original $T=1$	6	7	7	8	7	6	7	6	6	8	6	6	7	6	6.6
Original $T=10$	8	9	7	9	8	6	8	9	7	8	7	7	6	7	7.6
Residual $T=1$	7	7	7	6	8	10	5	7	6	6	9	8	8	8	7.3
Residual $T=2$	7	6	6	7	7	18	5	6	8	6	10	8	8	8	7.9
Test case 3															
Original $T=1$	5	10	8	5	5	6	8	5	5	5	5	6			6.1
Original $T=15$	7	6	7	7	7	7	5	8	7	5	6	7			6.6
Residual $T=1$	19	6	11	13	6	7	6	8	12	17	10	9			10.3
Residual $T=3$	11	8	19	8	6	6	7	6	8	8	8	7			8.5

Table 4
The surrogate method results for test case 1 according to tests: DVV, Third order auto covariance (C3), time reversibility (REV), Standard deviation (SD), central frequency (CF), Correlation Dimension (CD) and LLE (L=Linear, N=Nonlinear).

Series	Wind speed series									Residual series								
	DVV	C3	REV	SD	CF	CD $D=10$ $T=1$	CD $D=10$ $T=5$	LLE $D=10$ $T=1$	LLE $D=10$ $T=5$	DVV	C3	REV	SD	CF	CD $D=9$ $T=1$	CD $D=9$ $T=3$	LLE $D=9$ $T=1$	LLE $D=9$ $T=3$
1	L	L	L	N	L	L	N	N	L	N	L	L	N	N	L	N	L	L
2	N	N	N	N	L	N	N	L	N	N	L	L	N	N	L	N	L	L
3	L	N	L	N	L	N	L	L	N	N	L	N	N	L	N	L	L	L
4	N	L	L	N	N	N	N	N	N	N	N	L	N	L	L	L	N	L
5	L	L	N	N	N	L	L	L	N	N	N	N	L	L	N	L	N	L
6	N	L	N	N	L	L	N	L	N	N	N	L	N	L	L	L	L	N
Sum of Nonlinear series	3	2	3	6	2	3	3	2	5	6	3	2	5	2	2	2	2	1

Table 5
The surrogate method results for test case 2 according to tests: DVV, Third order auto covariance (C3), time reversibility (REV), Standard deviation (SD), central frequency (CF), Correlation Dimension (CD) and LLE (L=Linear, N=Nonlinear).

Series	Wind speed series									Residual series								
	DVV	C3	REV	SD	CF	CD $D=7$ $T=1$	CD $D=8$ $T=10$	LLE $D=7$ $T=1$	LLE $D=8$ $T=10$	DVV	C3	REV	SD	CF	CD $D=7$ $T=1$	CD $D=8$ $T=2$	LLE $D=7$ $T=1$	LLE $D=8$ $T=2$
1	N	L	N	L	N	N	N	L	L	N	L	N	N	N	L	L	N	N
2	N	N	L	L	N	N	L	N	L	N	L	N	N	L	N	L	N	L
3	N	N	L	L	N	N	L	L	L	N	L	L	N	N	N	N	L	L
4	N	N	N	N	L	N	N	L	L	N	N	N	N	L	N	L	N	N
5	N	N	N	N	L	N	N	N	L	N	L	N	N	L	L	L	L	N
6	N	N	N	N	L	N	N	L	L	N	N	L	N	L	N	N	L	N
7	N	N	N	N	L	N	N	L	N	N	N	N	N	L	N	N	N	L
8	N	N	N	N	N	N	N	L	L	N	L	L	N	L	N	L	N	N
9	L	N	N	L	L	N	L	N	L	N	N	N	N	L	L	L	N	L
10	N	N	N	L	N	N	N	N	L	N	L	N	N	L	N	L	L	L
11	N	N	L	L	L	N	N	L	L	N	N	N	N	L	N	N	N	N
12	N	N	N	N	L	N	N	L	N	N	N	N	N	L	N	N	L	L
13	N	N	L	L	L	N	N	L	L	N	L	N	N	L	N	L	N	N
14	N	N	N	L	L	N	N	N	N	N	L	N	N	L	N	N	L	N
Sum of Nonlinear series	13	13	10	6	5	14	11	5	3	14	6	11	14	2	11	6	8	8

Table 6
The surrogate method results for test case 3 according to tests: DVV, Third order auto covariance (C3), time reversibility (REV), Standard deviation (SD), central frequency (CF), Correlation Dimension (CD) and LLE (L=Linear, N=Nonlinear).

	Wind speed series									Residual series								
	DVV	C3	REV	SD	CF	CD $D=6$ $T=1$	CD $D=7$ $T=15$	LLE $D=6$ $T=1$	LLE $D=7$ $T=15$	DVV	C3	REV	SD	CF	CD $D=10$ $T=1$	CD $D=9$ $T=3$	LLE $D=10$ $T=1$	LLE $D=9$ $T=3$
1	N	L	N	N	N	N	N	L	N	N	L	N	N	N	L	L	N	L
2	N	L	N	N	L	N	N	N	N	N	N	N	N	L	L	L	N	N
3	N	N	L	N	L	N	N	N	N	N	L	N	N	L	L	N	N	N
4	N	N	L	N	L	N	N	L	L	N	L	N	N	L	L	L	N	N
5	L	L	N	N	L	N	N	N	L	N	N	N	N	L	L	L	N	N
6	N	N	N	L	L	L	N	L	N	N	N	N	N	L	L	N	N	N
7	N	N	N	N	L	N	N	N	L	N	N	N	N	L	L	L	N	N
8	N	N	N	N	L	N	N	N	N	N	N	N	N	N	L	L	N	N
9	N	L	N	N	L	N	N	N	N	N	L	N	N	L	L	L	N	N
10	N	L	N	L	L	L	N	L	L	N	N	N	N	L	L	L	N	N
11	L	L	N	N	L	L	N	L	N	N	N	N	N	L	L	N	L	L
12	N	L	N	N	N	L	N	L	L	N	L	N	N	L	L	N	N	N
Sum of Nonlinear series	10	5	10	10	2	8	12	6	7	12	7	12	12	2	0	4	11	2

measures confirm smaller number of nonlinear signals among residuals in comparison with original wind speed time series. High values of unpredictable to deterministic components ratio explains the obtained results. In other words, in these methods the ND component is negligible in the presence of the high unpredictable components. As a result, with these methods the wind speed signals are represented linearly because of the linear unpredictable components. Consequently, such measures are appropriate for signals whose noise to deterministic components ratio is small.

5.5. *Delay vector variance method*

DVV is a well-known technique to investigate the existence of the deterministic and nondeterministic components in a time series. Moreover, it is utilized as nonlinear measure in the surrogate method to determine the nonlinear characteristic of a process [67].

This technique is based on the time delay embedding representation of a time series. Time delay embedding can be utilized in order to represent a time series in the phase space. The DVs of a time series $x(n)$ of a given embedding dimension d can be denoted as $y_k=[x_{k-dT}, x_{k-(d-1)T}, \ldots, x_{k-T}]$, in which T is the time lag. This vector is comprised of d consecutive time samples. A corresponding target namely following sample $x(k)$ is assigned to each DV. The following paragraphs describe the DVV method procedure.

The DVs whose Euclidian distance l to the DV $y(k)$ lays within a certain distance should be first identified and then grouped together into a set $B_k(l, d)$. In order to investigate the complete range of pairwise distances, the threshold is automatically scaled with the dynamical range of the given time series as well as the embedding dimension. The variation of this distance is standardized according to the distribution of pair wise distance between DVs.

In the next stage, the standard deviation σ_d and the mean μ_d are calculated for all pair wise Euclidean distances between all DVs. All the DVVs whose distance to $y(k)$ is smaller than $l \in [\mu_d - n_d\sigma_d; \mu_d + n_d\sigma_d]$ are included in a set $B_k(l,d)=\{y_i \,|\, \|y_k - y_i\| \le l\}$. The span over which the DVV analysis is performed is controlled by n_d. For every B_k variance of the corresponding target is calculated and used to normalize the average overall sets B_k

$$\sigma^{*2}=\frac{1}{N\sigma_x^2}\sum_{k=1}^{N}\sigma_k^2 \qquad (21)$$

In which σ^{*2} is the inverse measure of the predictability. The standardization of the distance axis results in easy interpretation of the DVV plots. The small target variances stem from strong deterministic components. At the extreme right, the variance of the time series is the same value as the variance of the targets. Therefore, the DVV plots smoothly converge to unity due to occurrence of the maximum spans.

In order to determine the linear or nonlinear characteristic of a time series, one should investigate the DVV plot of the original time series as well as a number of its surrogate time series. The distance axis is standardized. Hence, a scatter diagram can be used to combine these plots. In this diagram, the vertical axis is associated with the DVV plot of the surrogate time series and the horizontal one is corresponded to the DVV plot of the original time series. A time series is considered to be linear if its surrogate time series DVV plots are similar to those of the original one. The nonlinearity of a time series is approved if its DVV plot is significantly different comparing with its surrogate DVV plots. Fig. 2 shows the DVV plots for the residual and original time series for the first series of test case 1. The solid curve represents the original signal and the dashed curve shows the mean of DVV plots for all surrogates.

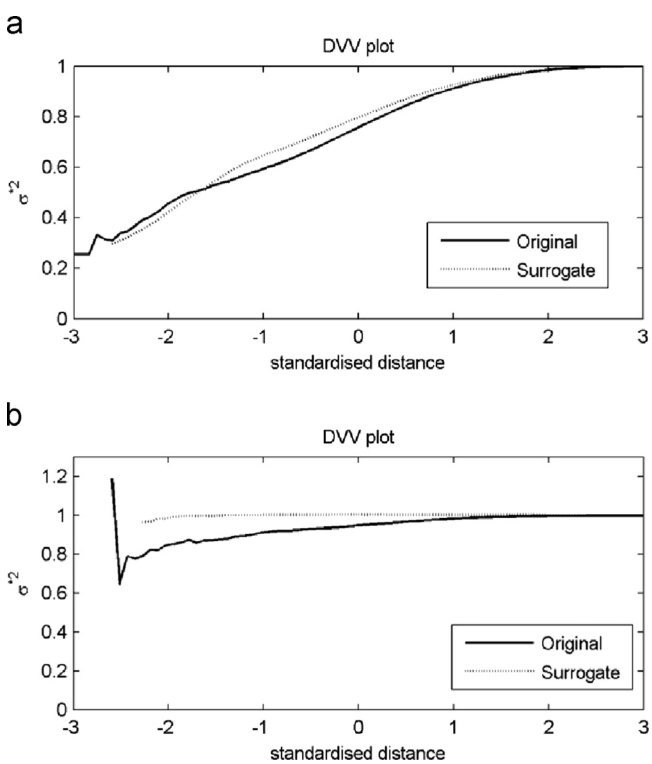

Fig. 2. DVV plot of signals related to the (a) original and (b) residual for the first series of test case 1.

In [67], a nonlinear measure is proposed in order to examine the nonlinearity of a time series. This measure is based on quantification of the difference between DVV plots of the original time series and its surrogates. It can be employed as a nonlinearity test in surrogate data method. The second column of Tables 4–6 shows the results of the DVV nonlinearity test on the original and residual wind time series for the three test cases. The DVV method considers only the deterministic components of signals, so it is insensitive to stochastic components and leads to more realistic results comparing to those of the previously discussed methods. The results in Tables 4–6 show that all residual signals are nonlinear.

6. The proposed quantizing of the nonlinear deterministic component

In this study, a quantization approach is proposed instead of strict answer about the nonlinearity of time series. It is illustrated in Fig. 2 that for the original signal σ^2_{min} is about 0.23 which is smaller than that of residual signal 0.61. This affirms that the deterministic components of the original signal are much more than that of the residual signals. In this study, in order to determine the nonlinear nature of a time series, an index which is a ratio of the nonlinear to NR is employed. This index is mathematically expressed in Eq. (3). The value of this index lays within the range of 0–1. If the ND component is low in comparison to TD component, NR tends to zero and vice versa.

The second row of Tables 7–9 shows the NR for all years. The average value of NRs for the first to third test cases is 0.4335, 0.1965 and 0.0135 respectively. These values demonstrate the considerable power of nonlinear component in the wind speed time series (test cases 1 and 2). This assesses that employing individual linear models for forecasting the wind velocity cannot be sufficient because of the strong presence of non-linear deterministic components. Hence, in order to provide precise prediction one should use the nonlinear methods beside linear ones, which is demonstrated in the next section. In other hand a negligible value for ND component is observed for the third case which is related to the out power of wind turbines which means the considerable reduction of ND component after transforming the wind speed to wind power. However in the next section we show with using nonlinear prediction methods this small value can be even further reduced.

The set of Tables 3–6 with set of Tables 7–9 provide a comparison between the results of applying various tests of the surrogate method (9 tests) and the proposed method. For example for results corresponding to first series of test case 1, the result of 3 tests is nonlinear and 6 tests result the linear model. Hence the

Table 7
NR of test case 1.

Series	1	2	3	4	5	6	Mean
NR	0.448	0.4518	0.3999	0.4471	0.4368	0.4172	0.4335
NR Markov	0.2276	0.2767	0.1971	0.1932	0.2434	0.2307	0.2281
% usage of ND with Markov	49.2	38.75	50.71	56.79	44.29	44.69	47.4
NR Grey	0.2595	0.2699	0.2533	0.2755	0.3338	0.3116	0.2839
% usage of ND with Grey	42.08	40.27	36.65	38.38	23.57	25.3	34.38
NR Grey–Markov	0.1613	0.1782	0.1609	0.182	0.2091	0.1674	0.1765
% usage of ND with Grey–Markov	64	60.56	59.77	59.3	52.14	59.88	59.28
NR EMD–Grey	0.2544	0.2669	0.2292	0.2473	0.247	0.2295	0.2457
% usage of ND with EMD-Grey	43.22	40.91	42.7	44.68	43.45	45	43.33
NR NARMA	0.44108	0.44298	0.38351	0.39469	0.432	0.34293	0.4062
% usage of ND with NARMA	1.55	1.95	4.1	11.72	1.1	17.8	6.37

Table 8
NR of test case 2.

Series	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Mean
NR	0.2601	0.4059	0.2937	0.2404	0.1975	0.0981	0.1191	0.161	0.1969	0.1635	0.1628	0.1426	0.1926	0.117	0.1965
NR Markov	0.1411	0.2407	0.1659	0.1636	0.0896	0.0485	0.0488	0.0792	0.129	0.1119	0.1067	0.0862	0.0926	0.0617	0.1118
% usage of ND with Markov	45.74	40.69	43.52	31.97	54.64	50.54	59.03	50.81	34.49	31.54	34.46	39.53	51.94	47.31	44.01
NR Grey	0.1689	0.2897	0.1766	0.1433	0.1925	0.084	0.0964	0.0964	0.1199	0.0989	0.0932	0.1238	0.1183	0.0805	0.1345
% usage of ND with Grey	35.08	28.62	39.86	40.41	2.52	14.34	19.08	40.11	39.11	39.5	42.73	13.18	38.59	31.17	30.31
NR Grey–Markov	0.0849	0.2057	0.1217	0.0988	0.089	0.0464	0.0535	0.057	0.0801	0.075	0.069	0.0745	0.0893	0.0601	0.0861
% usage of ND with Grey–Markov	67.36	49.32	58.56	58.9	54.92	52.66	55.11	64.61	59.33	54.11	57.62	47.74	53.62	48.64	55.89
NR EMD–Grey	0.1541	0.2364	0.1718	0.1431	0.1262	0.0593	0.0822	0.1043	0.1158	0.097	0.0906	0.0856	0.1215	0.0771	0.1189
% usage of ND with EMD–Grey	40.76	41.76	41.49	40.46	36.11	39.6	30.95	35.2	41.2	40.68	44.35	39.97	36.9	34.08	38.82
NR NARMA	0.2449	0.3388	0.238	0.2176	0.1809	0.1093	0.1315	0.1489	0.1842	0.1568	0.1496	0.1275	0.2082	0.0832	0.18
% usage of ND with NARMA	5.85	16.53	18.97	9.46	8.4	−11.38	−10.37	7.53	6.44	4.08	8.09	10.6	−8.12	28.86	6.78
NR ARMAX	0.2525	0.2907	0.2894	0.2503	0.1847	0.0955	0.1045	0.1577	0.161	0.2154	0.1606	0.1497	0.1824	0.0926	0.1848
% usage of ND with ARMAX	2.92	28.39	1.45	−4.1	6.47	2.62	12.23	2.07	18.23	−31.73	1.37	−4.97	5.28	20.88	4.37

Table 9
NR of test case 3.

Series	1	2	3	4	5	6	7	8	9	10	11	12	Mean
NR	0.017	0.0179	0.0189	0.0163	0.0116	0.0047	0.0107	0.0119	0.0051	0.0082	0.0218	0.0177	0.0135
NR Markov	0.0037	0.0035	0.0044	0.0045	0.0041	0.0018	0.0057	0.0095	0.0014	0.0032	0.0075	0.0062	0.0046
% usage of ND with Markov	77.93	80.4	76.59	72.5	64.87	61.37	46.35	20.05	72.75	61.67	65.65	65.04	63.76
NR Grey	0.0106	0.0189	0.0062	0.0135	0.0059	0.0028	0.0072	0.0055	0.0039	0.0084	0.0076	0.0184	0.0091
% usage of ND with Grey	37.36	−5.66	66.95	17.21	49.5	40.26	32.87	53.96	23.05	−2.35	65.04	−3.93	31.19
NR Grey–Markov	0.0009	0.0013	0.0022	0.0026	0.0013	0.0008	0.0021	0.0022	0.0008	0.003	0.0033	0.0028	0.0019
% usage of ND with Grey–Markov	94.93	92.55	88.47	83.88	89.19	83.37	80.06	81.49	83.54	63.62	84.95	84.26	84.19
NR EMD–Grey	0.0121	0.0105	0.0123	0.0152	0.005	0.0025	0.0063	0.008	0.0044	0.0053	0.0191	0.0059	0.0089
% usage of ND with EMD–Grey	28.73	41.52	34.95	6.98	56.68	46.87	41.34	32.89	14.22	35.31	12.3	66.75	34.88
NR NARMA	0.01	0.0146	0.0158	0.0156	0.0116	0.0037	0.0051	0.0049	0.0045	0.0081	0.0209	0.0192	0.0112
% usage of ND with NARMA	40.96	18.25	16.48	4.05	0	20.79	52.16	59.02	12.12	1.75	4.2	−8.46	18.41

results of various tests are in conflict with each other. Instead of such strict answers which are in conflict with each other our proposed index is a number equal to 0.448 which is the ratio of ND to TD components. Another example is for the second series of test case 1 for which 7 tests confirm that it is nonlinear and 2 tests show it is linear. However, the corresponding index is 0.4675. Therefore, it is obvious the idea of strict answer for nonlinearity is not clear enough and instead using an index that gives us a number between 0 and 1 can be a better alternative idea.

7. Evaluating some prediction methods with aid of the proposed index

After applying linear ARMA models on the original series, nonlinear models can also be employed on the residual series to use their ND components for prediction. In other words, we aim to predict the residual series which contains ND and U components ($h(.)+e_k$ in Eq. (2)). The employed prediction methods and their results are described in following subsections. So it should note the input of the following methods is the residual series.

7.1. Markov chain

Here Markov chain method is employed to predict the residual series. The theory and details of using this method for prediction is given in [44]. In this study a data window includes previous 70 samples with states number equal to 50 are considered to build probability transition matrix. Probability transition matrices with orders from 1 to 3 are considered.

The third row of Tables 7–9 gives the NR for all series of the three test cases. The values show the large reduction of NR after using Markov. For example according to Table 7 the average of NR of the six series for test case 1 is reduced from 0.4335 to 0.2281. The fourth row in Tables 7–9 show the percentage reducing of NR after using Markov method for each series. For example for the first test case the average of this value over 6 series is 47.4%. This means Markov method has the ability of pulling 47.4% of nonlinear deterministic component for the prediction. The average of percentage reducing for test cases 2 and 3 is

44.01% and 63.76% respectively. All these numbers approve that the Markov chain method can be effective in pulling the ND component of wind speed and wind power time series.

7.2. Grey model

In this section Grey model is used to predict the residual series. Basically, in the grey system theory GM (n,m) denotes for grey model. In this case, n stands for the order of the difference equation and m shows the number of variables. Here GM (1,1) is used due to the fact that it is fast and require a small amount of computation effort and at the same time it has acceptable accuracy. The detail for Grey models is given in [77,78]. Conventionally, when the number of samples is too large, rolling Grey system will be employed which requires less computation effort. In this method the number of samples used for prediction remains constant and when the latest sample enters, the last sample leaves the prediction window. This method is used in this study due to the large number of samples. The number of samples employed in each window is 5 which results in acceptable accuracy with reasonable computation effort.

The fifth and sixth rows of Tables 7–9 show the results of applying GM (1,1) on the residuals. The results show the percentage reduction of NR is 34.38%, 30.31% and 31.19% for test cases 1–3 respectively. These values show Grey model is also effective in using ND component for prediction. But however it is weaker than Markov.

7.3. Grey–Markov

In this method firstly, Grey model is used for prediction the residual series. Then the prediction error is fed to Markov chain method to be predicted and further reducing the total prediction error. The results are shown in rows 7 and 8 in Tables 7–9. The reduction of NR is 59.28%, 55.89% and 84.19% for test cases 1–3 respectively. These values show the superiority of this method over all other considered methods in our study.

7.4. EMD–Grey

The EMD technique is introduced in [79]. Such algorithm is based on the interpolation point's criterion that decomposes them into a number of amplitude and frequency modulated with zero mean, called Intrinsic Mode Functions (IMFs) [79,80].

In EMD–Grey method in the first step by use of EMD the residual series is decomposed to some IMFs. In the second step each IMF is predicted by Grey model. In last step the final prediction is attained by sum of all IMFs predicted values.

The results of this method are shown in rows 9 and 10 in Tables 7–9. The values show reductions of NR equal to 43.3%, 38.82% and 34.88% for test cases 1–3. In all test cases the results of EMD–Grey show a considerable improvement over the Grey method.

7.5. NAR network

Nonlinear Autoregressive Neural Network [81] is a nonlinear learning-based algorithm. This approach is a nonlinear generalization of the Box and Jenkins method. A multilayer perceptron and delayed samples of the original time series are used in this methodology to obtain the model parameters. In this study, the structure of the network is feed forward including three layers: input layer, hidden layer, and output layer. The training algorithm is Levenberg Marquardt and forty neurons are used in the hidden layer. The number of training target time steps is equal to 70% of the input length.

The results are shown in rows 11and 12 of Tables 7–9. The reduction of NR is equal to 6.37%, 6.78% and 18.41% for test cases 1–3 respectively. These values show that this method is weaker than previous methods for pulling the ND component. Also for some series the percentage reduction of NR becomes negative. It means in those special cases NARnet can not to pull the ND component from the residuals and employing it will increase the complexity of residual series.

7.6. ARMAX

In this section for test case 2 beside to present and past values of wind speed we also use the humidity and temperature to predict the wind speed. So it will be an ARMAX model. The ARMA model orders shown in Table 1 with using 4 recent samples of humidity and temperature are used for prediction. The results are shown in last two rows of Table 8. Results show average percentage reduction equal to 4.37% which the lowest among all discussed methods. Also there are some negative values of percentage reduction of NR for some series which mean the method was unable to use ND component for prediction in the related series.

8. Conclusions

In this paper a novel procedure is proposed to make judgment about the nonlinearity of time series. Instead of the strict answer for this question whether the time series is linear or nonlinear we give our answer as a number between 0 and 1. The results obtained from the various tests of surrogate method conflict with each other and fail in giving a clear answer about the nonlinearity of wind speed time series. The reason is the strictness of the answer of these tests. Our proposed method solves the problem with quantizing the nonlinear deterministic components. The performance of some prediction methods like Markov, Grey, Grey–Markov, EMD–Grey, NARnet and ARMAX is compared by using the proposed index. Results show Grey–Markov method has the best performance and can use the most of nonlinear deterministic component for prediction. On the other hand NARnet and ARMAX methods have the weakest performance.

References

[1] Joselin Herbert GM, Iniyan S, Sreevalsan E, Rajapandia S. A review of wind energy technologies. Renew Sustain Energy Rev 2007;11(6):1117–45.
[2] Soman SS, Zareipour H, Malik O, Mandal P. A review of wind power and wind speed forecasting methods with different time horizons. In: Proceedings of the North American Power Symposium (NAPS); 2010. p. 1–8.
[3] Lydia M, Kumar SS. A comprehensive overview on wind power forecasting. Singapore: IPEC; 2010; 268–73.
[4] McCarthy E. Wind speed forecasting in the central California wind resource area, EPRI-DOE-NREL Wind Energy Forecasting Meeting, Burlingame, CA; March 23, 1998.
[5] Centre Bureau of Meteorology Training. Meteorologist course: numerical weather prediction. Melbourne: Bureau of Meteorology Training Centre; 1994. p. 1096848.
[6] Al-Yahyai S, Charabi Y, Gastli A. Review of the use of Numerical Weather Prediction (NWP) models for wind energy assessment. Renew Sustain Energy Rev 2010;14(9):3192–8.
[7] Chandra DR, Kumari MS, Sydulu M. A detailed literature review on wind forecasting. Sri Rangalatchum Dindigul: ICPEC; 2013; 630–4.
[8] Potter CW, Negnevitsky M. Very short-term wind forecasting for Tasmanian power generation. Power Syst 2006;21(2):965–72.
[9] Lange M, Focken U. New developments in wind energy forecasting. In: IEEE power and energy society general meeting 2008 – conversion and delivery of electrical energy in the 21st century. p. 1-8, 20–4.
[10] Karki R, Thapa S, Billinton R. A simplified risk-based method for short-term wind power commitment. IEEE Trans on Sustain Energy 2012;3(3):498–505.
[11] Sideratos G, Hatziargyriou ND. An advanced statistical method for wind power forecasting. IEEE Trans Power Syst 2007;22(1):258–65.

[12] Jiakun A, Zhaohong B, Xiaozhong C, Bowen H, Shiyu L. A generalized data preprocessing method for wind power prediction. Vancouver: IEEE Power Energy Soc Gen Meet (PES); 2013; 1–5.
[13] Carta JA, Ramírez P, Velázquez S. A review of wind speed probability distributions used in wind energy analysis: case studies in the Canary Islands. Renew Sustain Energy Rev 2009;13(5):933–55.
[14] Oner Y, Ozcira S, Bekiroglu N, Senol I. A comparative analysis of wind power density prediction methods for ÇanakkaleIntepe region, Turkey. Renew Sustain Energy Rev 2013;23:491–502.
[15] Kusiak A, Zijun Z. Short-horizon prediction of wind power: a data-driven approach. EEE Trans Energy Convers 2010;25(4) (1112, 1122).
[16] Milligan M, Schwartz M, Wan Y. Statistical wind power forecasting models: results for U.S. wind farms, presented at WINDPOWER 2003, Austin, Texas; May 2003.
[17] Rajagopalan S, Santoso S. Wind power forecasting and error analysis using the autoregressive moving average modeling. Calgary: IEEE PES; 2009; 1–6.
[18] Chen P, Pedersen T, Bak-Jensen B, Chen Z. ARIMA-based time series model of stochastic wind power generation. IEEE Trans Power Syst 2010;25(2):667–76.
[19] Rajesh GK, Seetharaman K. Day-ahead wind speed forecasting using f-ARIMA models. J Renew Energy 2009;34:1388–93.
[20] Costa A, Crespo A, Navarro J, Lizcano G, Madsen H, Feitosa E. A review on the young history of the wind power short-term prediction. Renew Sustain Energy Rev 2008;12(6):1725–44.
[21] Chengwei T, Lei D, Shuang G, Xiaozhong L. A study of grey theory used in prediction of annual wind power generation. Wuhan: Int Conf Electr Inf Control Eng (ICEICE); 2011; 1952–5.
[22] Huang CY, Liu YW, Tzeng WC, Wang PY. Short term wind speed predictions by using the grey prediction model based forecast method. Baton Rouge: Green Technol. Conf. (IEEE-Green); 2011; 1–5.
[23] Zhang F, Cho K, Choi J, Lee YM, Lee KY. A study on wind speed prediction using artificial neural network at Jeju Island in Korea II. In: Proceedings of the IEEE 54th international midwest symposium circuits and systems (MWSCAS); 2011. p. 1–4.
[24] Ghanbarzadeh A, Noghrehabadi AR, Behrang MA, Assareh E. Wind speed prediction based on simple meteorological data using artificial neural network. In: Proceedings of the 7th IEEE international conference on industrial informatics; 2009. p. 664–7.
[25] de Aquino RRB, Gouveia HTV, Lira MMS, Ferreira AA, Neto ON, Carvalho MA. Wind forecasting and wind power generation: looking for the best model based on artificial intelligence. In: Proceedings of the international joint conference on neural networks (IJCNN); 2012. p. 1–8.
[26] Carolin Mabel M, Fernandez E. Analysis of wind power generation and prediction using ANN: a case study. J Renew Energy 2008;33:986–92.
[27] Steftos A. A novel approach for the forecasting of mean hourly wind speed time series. J Renew Energy 2002;27:163–74.
[28] Salcedo-Sanz S, Bellido, Garcia, Figuras, Prieto L, Correoso F. Accurate short term wind speed prediction by exploiting diversity in input data using banks of artificial neural networks. J Neurocomput 2009;71:1336–41.
[29] Lin KP, Huang XY, Liang WL, Chen BL. A nonlinear ensemble prediction model based on genetic for calculation wind of the wind assessment. In: Proceedings of the third international joint conference on computational science and optimization (CSO). vol. 2; 2010. p. 471–5.
[30] Xingpei L, Yibing L, Weidong X. Wind speed prediction based on genetic neural network. In: Proceedings of the 4th IEEE conference on industrial electronics and applications (ICIEA); 2009. p. 2448–51.
[31] Guo S, Li Y, Xiao S. Wind speed forecasting of genetic neural model based on rough set theory. Crit Infrastruct (CRIS) 2010:1–6.
[32] Xia J, Zhao P, Dai Y. Neuro-fuzzy networks for short-term wind power forecasting. Hangzhou: Int Conf Power Syst Technol (POWERCON); 2010; 1–5.
[33] Ghiasvand O., Ghiasvand A. Wind speed short term forecast by neuro fuzzy modeling with aid of Mutual Information at Manjil Wind Power Plant. In: Proceedings of the 1st international econference on computer and knowledge engineering (ICCKE); 2011. p. 1–5.
[34] Damousis I, Alexiadis MC, Theocharis JB, Dokopoulos PS. A fuzzy model for wind speed prediction and power generation in wind parks using spatial correlation. IEEE Trans Energy Convers 2004;19:352–60.
[35] Barbounis TG, Theocharis JB. A locally recurrent fuzzy neural network with application to wind speed prediction using spatial correlation. J Neurocomput 2007;70:1525–42.
[36] Damousis IG, Dokopoulos P. A fuzzy expert system for the forecasting of wind speed and power generation in wind farms. Sydney: ICPICA; 2001; 63–9.
[37] Karki R, Thapa S, Billinton R. A simplified risk-based method for short-term wind power commitment. IEEE Trans Sustain Energy 2012;3(3):498–505.
[38] Jianwu Z, Wei Q. Support vector machine -based short-term wind power forecasting. Phoenix: Power Syst Conf Expos (PSCE); 2011; 1–8.
[39] Yan J, Liu Y, Han S, Qiu M. Wind power grouping forecasts and its uncertainty analysis using optimized relevance vector machine. Renew and Sustain Energy Rev 2013;27:613–21.
[40] Sreelakshmi, Kumar R. Performance evaluation of short-term wind speed prediction techniques. Int J Comput Sci Netw Secur 2008;8(8):162–9.
[41] Negnevitsky M, Mandal P, Srivastava AK. Machine learning applications for load, price and wind power prediction in power systems. In: Proceedings of the 15th international conference on intelligent system applications to power systems; 2009. p. 1–6.
[42] Haque AU, Mandal P, Kaye ME, Meng J, Chang L, Senjyu T. A new strategy for predicting short-term wind speed using soft computing models. Renew Sustain Energy Rev 2012;16(7):4563–73.
[43] Yanning Z, Longyun K, Shiqiong Z, Binggang C. Wind speed predicted by wavelet analysis in input prediction control of wind turbine. In: Proceedings of the 7th world congress on intelligent control and automation; 2008. p. 7732–35.
[44] Murugesan S, Zhang J, Vittal V. Finite state Markov chain model for wind generation forecast: a data-driven spatiotemporal approach. Washingotn: Innov Smart Grid Technol (ISGT); 2012; 1–8.
[45] Lopes VV, Scholz T, Estanqueiro A, Novais AQ. On the use of Markov chain models for the analysis of wind power time-series. In: Proceedings of the 11th international conference on environment and electrical engineering (EEEIC); 2012. p. 770–5.
[46] Carpinone A, Langella R, Testa A, Giorgio M. Very short-term probabilistic wind power forecasting based on Markov chain models. In: Proceedings of the EEE 11th international conference on probabilistic methods applied to power systems (PMAPS); 2010. p. 107–12.
[47] Abdel-Karim N, Small M, Ilic M. Short term wind speed prediction by finite and infinite impulse response filters: a state space model representation using discrete markov process. Bucharest: IEEE Bucharest PowerTech; 2009; 1–8.
[48] Alexiadis MC, Dokopoulos PS, Sahsamanoglou HS. Wind speed and power forecasting based on spatial correlation models. IEEE Trans Energy Convers 1999;14(3):836–42.
[49] Hur J, Baldick R. Spatial prediction of wind farm outputs using the Augmented Kriging-based Model. San Diego: IEEE Power Energy Soc Gen Meet; 2012; 1–7.
[50] Barbounis TG, Theocharis JB. Locally recurrent neural networks optimal filtering algorithms: application to wind speed prediction using spatial correlation. IEEE Int Joint Conf Neural Netw 2005;5:2711–6.
[51] Gang L, Jinfu C, Defu C, Dongyuan S, Xianzhong D. Probabilistic assessment of available transfer capability considering spatial correlation in wind power integrated system. IET Gener Transm Distrib 2013;7(12):1527–35.
[52] Lee D, Baldick R. Short-term wind power ensemble prediction based on gaussian processes and neural networks. IEEE Trans Smart Grid 2014;5 (1):501–10.
[53] Han S, Liu Y, Yan J. Neural network ensemble method study for wind power prediction. Power and Energy Eng Conf (APPEEC) 2011:1–4 (25–28).
[54] Yongqian L, Shi J, Yang Y, Lee WJ. Short-term wind-power prediction based on wavelet transform – support vector and statistic-characteristics analysis. IEEE Trans Ind Appl 2012;48(4):1136–41.
[55] Chuanan Yao, Yongchang Yu. A hybrid model to forecast wind speed based on wavelet and neural network. International Conference on Control, Automation and Systems Engineering (CASE); 2011. p. 1–4.
[56] Pousinho HMI, Catalao JPS, Mendes VMF. Wind power short-term prediction by a hybrid PSO-ANFIS approach. In: Proceedings of the 15th IEEE Mediterranean electrotechnical conference (MELECON); 2010. p. 955–60.
[57] Catalao JPS, Pousinho HMI, Mendes VMF. Hybrid wavelet-PSO-ANFIS approach for short-term wind power forecasting in Portugal. IEEE Power and Energy Society General Meeting; 2011.
[58] Sharma D, Lie TT. Wind speed forecasting using hybrid ANN–Kalman filter techniques. Ho Chi Minh City: Conf Power Energy (IPEC); 2012; 644–8.
[59] Amjady N, Keynia F, Zareipour H. Wind power prediction by a new forecast engine composed of modified hybrid neural network and enhanced particle swarm optimization. IEEE Trans Sustain Energy 2011;2(3):265–76.
[60] Sheng C, Lin Y, Gengwu Z, Cheng Z, Shijun D, Chao D. Short-term wind power prediction based on combined Grey–Markov model. 2011 Int Conf Adv Power Syst Autom Prot (APAP) 2011;3:1705–11.
[61] Lin WM, Hong CM, Huang CH, Ou TC. Hybrid control of a wind induction generator based on Grey–Elman neural network. IEEE Trans Control Syst Technol 2013;21(6):2367–73.
[62] Liu XJ, Mi ZQ, Lu B, Tao W. A Novel Approach for Wind Speed Forecasting Based on EMD and Time-Series Analysis. Wuhan: Asia-Pac Power Energy Eng Conf (APPEEC); 2009; 1–4.
[63] Guo Z, Zhao W, Lu H, Wang J. Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model. Renew Energy 2012;37 (1):241–9.
[64] Lutkepohl H. Introduction to multiple time series analysis. Berlin: Springer-Verlag Publ.; 1991; 767.
[65] Abarbanel HDI, Brown R, Kennel MB. Determining embedding dimension for phase-space reconstruction using a geometrical construction. Phys Rev A 1992;45:3403–11.
[66] Eubank S, Farmer JD, Galdrikian B, Longtin A, Theiler J. Testing for nonlinearity in time-series: the method of surrogate data. Physica D 1992;58:77–94.
[67] Gautama T, Mandic DP, Van Hulle MM. A novel method for determining the nature of time series. IEEE Trans Biomed Eng 2004;51(5):728–36.
[68] Wold H. A study in the analysis of stationary time series. Uppsala, Almqvist & Wiksell, 1938:1954 (2nd ed.).
[69] Wei WWS. Time series analysis. Redwood City, California: Addison-Wesley Publ.; 1990; 614.
[70] Parlitz U. Nonlinear time series analysis. In: Suykens JAK, Vandewalle J, editors. Nonlinear modeling – advanced black-box techniques. Boston: Kluwer Academic Publishers; 1998. p. 256.
[71] Abarbanel H. Analysis of observed chaotic data. New York: Springer; 1996; 272.
[72] Parlitz U, Wedekind I, Lauterborn W, Merkwirthttingen C. TSTOOL and User Manual, Ver. 1.11, DPI Go; 2001. Available: 〈http://www.physik3.gwdg.de/tstool〉.
[73] Cao L. Practical method for determining the minimum embedding dimension of a scalar time series. Physica D 1997;110:43–50.

[74] Schmitz A, Schreiber T. Improved surrogate data for nonlinearity tests. Phys Rev Lett 1996;77:635–8.
[75] Schmitz A, Schreiber T. Discrimination power of measures for nonlinearity in a time series. Phys Rev E 1997;55:5443–7.
[76] Hilborn RC. Chaos and nonlinear dynamics: an introduction for scientists and engineers. Oxford: Oxford University Press; 2001.
[77] Julong D. Introduction to Grey system theory. J Grey Syst 1989;1:1–24.
[78] Shimizu N, Ueno O, Komata C. Introduction of time series data analysis using grey system theory. In: Proceedings of the 2nd international conference on knowledge-based intelligent electronic systems; 1998. p. 67–72.
[79] Huang NE, Shen Z, Long SR. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc R Soc Lond A 1998;45(4):903–95.
[80] Cohen L. Time-frequency analysis. NJ, USA: Prentice-Hall, Upper Saddle River; 1995.
[81] Chow TWS, Leung CT. Nonlinear autoregressive integrated neural network model for short-term load forecasting. IEE Gener Transm Distrib 1996;143(5):500–6.